

31759-190545

Akihiro OKUMURA et al.

日 本 国 特 許 庁  
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2003年 1月30日

出 願 番 号

Application Number:

特願2003-021979

[ ST.10/C ]:

[ JP 2003-021979 ]

出 願 人

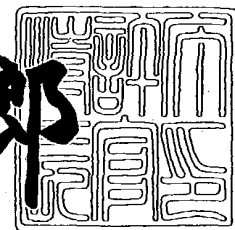
Applicant(s):

沖電気工業株式会社

2003年 4月 1日

特 許 庁 長 官  
Commissioner,  
Japan Patent Office

太田信一郎



出証番号 出証特2003-3022448

【書類名】 特許願

【整理番号】 KN002563

【提出日】 平成15年 1月30日

【あて先】 特許庁長官 太田 信一郎 殿

【国際特許分類】 G06F 15/00

【発明者】

    【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会  
社内

    【氏名】 松村 靖子

【発明者】

    【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会  
社内

    【氏名】 池野 篤司

【特許出願人】

    【識別番号】 000000295

    【氏名又は名称】 沖電気工業株式会社

    【代表者】 篠塚 勝正

【代理人】

    【識別番号】 100090620

    【弁理士】

    【氏名又は名称】 工藤 宣幸

【先の出願に基づく優先権主張】

    【出願番号】 特願2002-187625

    【出願日】 平成14年 6月27日

【手数料の表示】

    【予納台帳番号】 013664

    【納付金額】 21,000円

【提出物件の目録】

    【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9006358

【プールの要否】 要

【書類名】 明細書

【発明の名称】 情報処理装置および方法

【特許請求の範囲】

【請求項1】 複数の領域を含む所定の構造化文書の中から、主要な領域を選出する情報処理装置において、

前記構造化文書の表示構造を指定する制御文字を解析し、前記各領域に関する幾何学的な属性情報を生成する属性情報生成部と、

前記属性情報を比較することによって複数の領域の中から主要領域を選出する主要領域選出部とを備えたことを特徴とする情報処理装置。

【請求項2】 請求項1の情報処理装置において、

前記属性情報生成部は、前記属性情報として各領域の表示面積または表示面積の比率を用い、前記主要領域選出部は、当該表示面積または表示面積の比率が大きい領域を主要領域として選出することを特徴とする情報処理装置。

【請求項3】 請求項1の情報処理装置において、

前記属性情報生成部は、前記属性情報として各領域の表示位置を用い、前記主要領域選出部は、当該表示位置が表示画面上の中心に近い領域を主要領域として選出することを特徴とする情報処理装置。

【請求項4】 請求項1の情報処理装置において、

前記属性情報生成部は、前記属性情報として、各領域の表示面積または表示面積の比率、および各領域の表示位置を用い、

前記主要領域選出部は、当該表示面積もしくは表示面積の比率が大きい領域、または、表示位置が表示画面上の中心に近い領域を、主要領域として選出することを特徴とする情報処理装置。

【請求項5】 請求項4の情報処理装置において、

実数 $X$ 、 $Y$ 、 $\alpha$ 、 $\beta$ につき、 $X$ を前記各領域の表示面積または表示面積の比率、 $Y$ を前記表示画面上の中心と各領域の中心との距離、 $\alpha$ および $\beta$ を重み係数とすると、

前記主要領域選出部は、 $X\alpha + Y\beta$ の示す値が大きな領域を主要領域として選

出することを特徴とする情報処理装置。

【請求項 6】 請求項 4 の情報処理装置において、

実数 X、Y につき、X を前記各領域の表示面積または表示面積の比率、Y を前記表示画面上の中心と各領域の中心との距離とすると、

前記主要領域選出部は、X Y の示す値が大きな領域を主要領域として選出することを特徴とする情報処理装置。

【請求項 7】 請求項 2 の情報処理装置において、

前記属性情報生成部は、前記属性情報として各領域の表示面積または表示面積の比率を用いることの近似処理として、当該表示面積または表示面積の比率の替わりに、同じ文字種であって、なおかつ、同じ文字サイズの文字の計数結果を利用することを特徴とする情報処理装置。

【請求項 8】 複数の領域を含む所定の構造化文書の中から、主要な領域を選出する情報処理方法において、

属性情報生成部が、前記構造化文書の表示構造を指定する制御文字を解析して、前記各領域に関する幾何学的な属性情報を生成し、

主要領域選出部が、前記属性情報を比較することによって複数の領域の中から主要領域を選出することを特徴とする情報処理方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は情報処理装置および方法に関し、例えば、WWW (World Wide Web) サイトから構造化文書を取得する場合などに適用し得るものである。

【0002】

【従来の技術】

WWW サイトに存在する構造化文書を取得し閲覧するためのツールとして、WWW ブラウザがある。一般的に、構造化文書はその文書のページのレイアウト、文字の大きさなどを柔軟に指定することができるようになっている。特に、図 1 のように、タイトル（領域 A）、他の構造化文書へのリンク（領域 B）、本文（領域 C）、その他（領域 D や E）など、ページがいくつかの領域（フレーム）に

分割されて、WWWブラウザに表示されるような構造化文書（フレームページ）が多く見られる。

【0003】

WWWブラウザを用いて、このような構造化文書から必要な情報を得るためには、ユーザは、目的の構造化文書のURLを指定し、その文書がWWWブラウザ上に表示された後に、文書をスクロールしながら目視により検索したり（人手による検索）、あるいは文字列検索機能を利用するといった作業を行なう必要がある。例えば、図1の領域Cが、ユーザの必要とする文書であったとし、こういった構造化文書が多数ある場合には、そのユーザが必要とする情報のみを複数の構造化文書から自動的にスクラップし、1つの文書にまとめてユーザに提示することが、人手による作業を簡略化する上で望ましくなる。このようなWWW情報抽出システムが、次の特許文献1に示されている。

【0004】

【特許文献1】

特開平10-187753号公報

【0005】

【発明が解決しようとする課題】

しかしながら、上記におけるWWW情報抽出システムでは、ユーザが構造化文書中で自分が必要とするデータの開始箇所と終了箇所をあらかじめ手入力により指定することが必要である。このため、大量の構造化文書に対して実施するにはユーザの操作負担が大きく現実的ではなかった。

【0006】

前記開始箇所や終了箇所を特定するためのスクラップデータ特定情報の選び方によっては、領域分割の形（フレーム構造）が異なる複数種類のフレームページを対象とする場合などにはフレームページごとに前記開始箇所や終了箇所の指定を変更することが必要となる可能性が高く、そのためのユーザの操作負担は大きい。

【0007】

【課題を解決するための手段】

かかる課題を解決するために、第1の発明では、複数の領域を含む所定の構造化文書の中から、主要な領域を選出する情報処理装置において、前記構造化文書の表示構造を指定する制御文字を解析し、前記各領域に関する幾何学的な属性情報を生成する属性情報生成部と、前記属性情報を比較することによって複数の領域の中から主要領域を選出する主要領域選出部とを備えたことを特徴とする。

【0008】

また、第2の発明では、複数の領域を含む所定の構造化文書の中から、主要な領域を選出する情報処理方法において、属性情報生成部が、前記構造化文書の表示構造を指定する制御文字を解析して、前記各領域に関する幾何学的な属性情報を生成し、主要領域選出部が、前記属性情報を比較することによって複数の領域の中から主要領域を選出することを特徴とする。

【0009】

【発明の実施の形態】

(A) 実施形態

以下、本発明にかかる情報処理装置および方法の実施形態について説明する。

【0010】

(A-1) 第1の実施形態の構成

本実施形態で、主要領域を判定し抽出する機能を有する領域処理部25（図1参照）の機能は、通信機能を有するパソコンその他の情報処理装置によって実現され得、WWWサーバ側に配置すること等も可能であるが、ここでは、通信端末（クライアント）側に配置する場合を例に説明する。

【0011】

本実施形態にかかる通信システム10の全体構成例を図10に示す。

【0012】

図10において、当該通信システム10は、ネットワーク11と、通信端末12と、WWWサーバ13とを備えている。

【0013】

このうちネットワーク11は、LAN（ローカルエリアネットワーク）などであってもかまわないが、ここでは、インターネットであるものとする。

【0014】

WWWサーバ13は、通信端末12からの要求（HTTPリクエスト）を受信すると、その要求に応じた応答（HTTPレスポンス）としてWWWページを構成するファイルなどを返送する機能を持つサーバである。多くの場合、WWWサーバ13は、予め生成したWWWページ等を蓄積しておくためのデータベース（図示せず）、そのデータベースを直接管理するデータベースサーバを伴う。また、これらWWWサーバ13やデータベースサーバなどの周辺には、ルータやファイアウォールなどの各種のネットワーク機器やDNSサーバなどのサーバ類が配置されてWWWサイトを構成するのが普通である。

【0015】

通信端末12は上述した領域処理部25を備えた情報処理装置で、具体的には、ネットワーク機能を有するパーソナルコンピュータなどであってよい。本実施形態の構成上、当該通信端末12は、WWWページを閲覧するためのプログラムであるWWWブラウザB1（図11参照）を搭載していることが必要である。

【0016】

当該通信端末12の内部構成例を図11に示す。

【0017】

（A-1-1）通信端末の内部構成例

図11において、当該通信端末12は、通信部20と、制御部21と、操作部22と、記憶部23と、表示部24と、領域処理部25とを備えている。

【0018】

このうち通信部20は、前記ネットワーク11を介してWWWサーバ13と通信する機能を持つ部分である。

【0019】

制御部21は、ハードウェア的には当該通信端末12の中央処理装置（CPU）に相当する部分であり、ソフトウェア的には、オペレーティングシステム（OS）や上述したWWWブラウザB1などに相当する部分である。

【0020】

操作部22は当該通信端末12のユーザU1が操作して制御部21に指示を伝



えるための部分で、例えば、キーボードや、ポインティングデバイスなどを有する。

#### 【0021】

表示部24は例えば液晶ディスプレイなどの表示画面を有する部分である。ユーザU1がWWWページを閲覧する場合には、WWWブラウザB1がWWWページのタグを解釈し処理した結果として、WWWページの内容が当該表示部24に画面表示され、ユーザU1に閲覧され得る。このとき画面表示されるWWWページは、一例として、図1に示したDP1であってよい。DP1のようなフレームのページ（フレームページ）を表示するためには、WWWブラウザB1はフレームに対応したものであることを要する。フレームとは、図1の各領域A～Eに画面表示されている内容（コンテンツ）のことではなく、その内容を収容している枠のことを指す。

#### 【0022】

また、本実施形態において、「領域」は、当該コンテンツの意味で使用するともあるが、基本的には当該フレームと同義なものとして使用する。

#### 【0023】

通常、1つのWWWページは、基本となる1つのHTMLファイルと、必要に応じて1または複数の各種ファイル（画像ファイルなど）から構成されるが、DP1のようなフレームページは、これよりファイル数も多く、複雑な構造を有する。

#### 【0024】

すなわちフレームページは、フレームの数や大きさ等を含むそのWWWページ全体の構成（すなわち、フレーム構造）を規定するHTMLファイル（フレーム規定ファイル）と、内容として各フレームに配置される複数のHTMLファイルが最低限、必要であり、これに加えて、各HTMLファイルにリンクされた各種ファイル（画像ファイルなど）が適宜、追加されることになる。

#### 【0025】

したがって、簡単のために各種ファイルがなく、HTMLファイルだけから構成されるフレームページを仮定しても、図1に示したフレームページDP1は、

フレームページDP1全体のフレーム構造を規定するフレーム規定ファイルと、図1の例では5つ設けられている各フレームに配置される5つのHTMLファイルから構成される必要があるから、全部で6つのHTMLファイルを要する。

## 【0026】

通常のWWWページは、1つのHTMLファイルの内部でのみ構造化されているが、フレームページでは、各HTMLファイルの内部で構造化されているだけでなく、1つのフレームページに含まれる複数のHTMLファイルのあいだにも構造が存在する。

## 【0027】

なお、図1では、各領域A～Eのあいだの境界線（スクロールバーなども含む）L1～L4が表示されているが、実際のフレームページでは、視覚的な効果などとの関連で、意図的にこのような境界線が表示されないようにし、異なる領域間でまったく地の色が同じになるようにしたり、まったく切れ目のない連続的な背景模様を表示したりすることも多い。したがって、境界線の表示の有無は、領域分割（フレーム構造）の本質とは関係ない。

## 【0028】

フレーム構造、すなわち、1つの画面をいくつのフレームに分割し、各フレームの辺の長さの割合（この割合は、各フレームの面積に対応）をどのように設定するか等（境界線の表示、非表示も含む）は、前記フレーム規定ファイル（例えば、DP11）の記述によって決まる。

## 【0029】

フレーム規定ファイルには様々な記述が行われるが、一例として、前記フレームページDP1を表示するためのフレーム規定ファイルDP11は、図3に示すようなものであってよい。ただし図3には、HTMLファイル（フレーム規定ファイル）のヘッダ部分などは省略して、本実施形態の構成上、重要な部分のみを図示してある。前記領域処理部25へ入力されるデータは図3に示すデータ（フレーム規定ファイルDP11の主要部）であってよい。もちろん、必要に応じて、HTMLファイルDP11をそのまま領域処理部25へ供給し、領域処理部25内で、図3に示すデータを抽出するようにしてもよいことは当然である。

【0030】

当該フレームページDP1の閲覧を希望する場合、ユーザU1が通信端末12のWWWブラウザB1に入力するのは、フレーム規定ファイルDP11のURL（ここでは、URL11とする）である。したがってこのとき、当該フレーム規定ファイルDP11の返送を要求するHTTPリクエストが通信端末12からWWWサーバ13へ送信され、そのHTTPレスポンスとして、各種HTTPヘッダ（エンティティヘッダも含む）とともに、エンティティボディとして当該フレーム規定ファイルDP11が返送される。

【0031】

エンティティボディ、すなわち、HTMLファイルや画像ファイルなどのファイルの返送を要求する場合、HTTPリクエストはGETメソッドを使用するGETリクエストになる。

【0032】

図1に示すフレームページDP1を構成する6つのHTMLファイルのうち当該フレーム規定ファイルDP11以外のHTMLファイルを、DP12～DP16とする。そして、HTMLファイルDP12は前記領域Aに対応し、HTMLファイルDP13は前記領域Bに対応し、HTMLファイルDP14は前記領域Cに対応し、HTMLファイルDP15は前記領域Dに対応し、HTMLファイルDP16は前記領域Eに対応するものとする。

【0033】

この場合、これらHTMLファイルDP12～DP16は、当該フレーム規定ファイルDP11が通信端末12に受信されたあと、当該フレーム規定ファイルDP11中の記述（URL12～URL16）に基づいて、WWWブラウザB1が順次、自動的に送信する各HTTPリクエストに応じた各HTTPレスポンスとして、通信端末12に受信される。

【0034】

そして、これら6つのHTMLファイルを処理し、整形した結果として、例えば、図1に示すような画面が、前記表示部24に表示されることになる。

【0035】

ここで、HTMLファイルDP12のURLを前記URL12とし、HTMLファイルDP13のURLを前記URL13とし、HTMLファイルDP14のURLを前記URL14とし、HTMLファイルDP15のURLを前記URL15とし、HTMLファイルDP16のURLを前記URL16とすると、URL12は図3の行TG12中の「title. html」であり、URL13は行TG13中の「link. html」であり、URL14は行TG14中の「honbun. html」であり、URL15は行TG15中の「sonota1. html」であり、URL16は行TG16中の「sonota2. html」である。

## 【0036】

通常、1つのフレームページを構成する複数のHTMLファイルは、フレーム規定ファイルと同じWWWサーバ（ここでは、13）に置かれ、しかも同じフォルダ内に置かれるため、FQDNを含まないこのようなローカルなURL（ここでは、ファイル名だけから成るURL）によって指定することができる。

## 【0037】

必要ならば、1つのフレームページを構成する複数のHTMLファイルを別なWWWサーバに置くこともでき、その場合には、例えば、前記URL12～URL16の全部または一部が、FQDNを含むURLに変わる。なお、ユーザU1によって入力され得るフレーム規定ファイルDP11を指定する前記URL11は、FQDNを含むURLであることは当然である。

## 【0038】

このURL11は、ユーザU1が操作部22を用いて行う手作業によって入力され得るほか、例えば、前記特許文献1に記載されたオートパイロットツールのようなソフトウェアを利用し、予めユーザU1が設定した日時や時間間隔で自動的に入力させることができる。

## 【0039】

図1のようなメニュー（領域Bがメニュー）を有するフレームページの場合、メニュー領域中の各メニュー（例えば、「営業本部」や「総務」など）をユーザU1が操作（選択）したときに、そのメニューに埋め込まれたURLのリンク先

のファイルが読み込まれ表示されるフレームを、当該メニュー領域Bに対応する前記HTMLファイルDP13の記述(target (ターゲット) 属性の値に示したフレーム名) によって指定するようになっている。

## 【0040】

図1のような構成の場合、各メニューのtarget属性の値としては、領域DP14が配置されているフレームのフレーム名を指定することが多い。なお、フレーム名とは、フレームページ中の各フレームを識別するために付与される名称で、フレームページ中の各HTMLファイルのURLの直後に記述する。図3の例では、例えば、HTMLファイルDP12のURLである"title. html"の直後の位置PS1に、当該HTMLファイルDP12を配置するフレームのフレーム名が記述される。

## 【0041】

したがって、典型的な使用方法では、メニューを有するフレームページの場合、他のフレームの内容（ここでは、領域A、領域B、領域D、領域E）は同じでターゲットのフレームの内容（領域C）だけが置き換わった複数のWWWページが重層的に存在しているとみることができる。

## 【0042】

図11に示す前記記憶部23は、例えば、RAM（ランダムアクセスメモリ）などの揮発性記憶装置や、ハードディスクなどの不揮発性記憶装置によって構成される。

## 【0043】

通信端末12が前記WWWサーバ13からWWWページを構成する各ファイル（例えば、前記DP11やDP12など）を受信すると、それらのファイルは、当該記憶部23のハードディスク上に確保されたキャッシュ領域に一時的に蓄積される。

## 【0044】

当該キャッシュ領域は、通常、WWWブラウザB1の管理下に置かれ、WWWブラウザB1から自由にアクセスすることが可能である。

## 【0045】

当該キャッシュ領域におけるファイルの記憶は可能なかぎり維持されるが、キャッシュ領域の記憶容量には上限があるため、この上限を越えて新たなWWWページを閲覧し、新たなファイルを蓄積するときには、すでに記憶されているファイルを例えば古いものから順番に削除することによって、必要な記憶容量を確保することになる。

【 0 0 4 6 】

例えばURLの入力などにより、ユーザU1が操作部22からWWWページの閲覧を指示した場合、そのURLで指定されるWWWページに関連するファイルがキャッシュ領域に記憶されていれば、WWWブラウザB1は、インターネット11経由ではなく当該キャッシュ領域から当該ファイルを取得し、表示部24に当該ファイルの内容を画面表示する。これによりネットワーク11上の通信トラフィックや、WWWサーバ13側の負荷の増大を抑制し、ユーザU1から見た応答時間（指示を出してからWWWページが表示されるまでの時間）を短縮することができる。

【 0 0 4 7 】

前記制御部21に接続された領域処理部25は、主要領域を判定し抽出する機能を有する部分であるが、その内部構成は、例えば、図2に示すようになる。

【 0 0 4 8 】

当該領域処理部25は、WWWブラウザB1の機能の一部として設けることも可能であるが、本実施形態ではWWWブラウザB1と別個に当該領域処理部25を設けている。

【 0 0 4 9 】

主要領域とは、WWWページ（フレームページ）上の複数の領域のうちユーザ（ここでは、U1）にとって最も重要であると推定できる1つの領域のことである。本実施形態では、基本的に、最も面積の大きい領域が主要領域であるものと想定している。したがって、例えば、図1に示したフレームページの場合なら、領域A～Eのうち、ユーザU1にとって最も面積の大きい1つの領域（ここでは、領域C）が主要領域となる。

【 0 0 5 0 】

この方法によれば、メニューを有するフレームページの典型的な使用方法に対応して、ターゲットのフレームの内容（例えば、領域C）だけが置き換わった複数のフレームページに関しても、何ら複雑な処理を追加することなく、領域Cを主要領域とすることができる。

【0051】

(A-1-2) 領域処理部の内部構成例

図2において、当該領域処理部25は、端子100、103と、領域抽出部101と、最大領域判定部102とを備えている。

【0052】

このうち端子100は入力端子である。入力端子100はHTML言語で記述された前記HTMLファイル（例えば、DP11）の供給を受ける部分であるが、当該入力端子100に対する当該HTMLファイルの直接の供給元については、領域処理部25とWWWブラウザB1の関係などにより、様々な変形があり得る。

【0053】

例えば、領域処理部25がWWWブラウザB1と連携してWWWサーバ13が送信したHTMLファイルをWWWブラウザB1経由で受け取る場合には、当該供給元はWWWブラウザB1となるが、当該領域処理部25がWWWブラウザB1と別個の独立したHTTPクライアントであり、WWWブラウザB1を経由せずにHTMLファイルを受け取る場合には、当該供給元は制御部21内の前記OSなどとなり得る。

【0054】

当該入力端子100に接続されている領域抽出部101は、入力されたHTMLファイルの中にある、領域分割（フレーム構造）に関する制御文字を解析することにより、分割された領域を抽出し、それぞれの分割された領域の表示される面積を計算する部分である。ここで計算された面積は判定基礎情報として最大領域判定部102へ供給する。

【0055】

このような面積の計算は、基本的に、前記フレーム規定ファイルDP11の記

述だけに基づいて実行可能である。図3に示した行SQ1, SQ2, SQ3をもとに、図1の各領域A～Eのフレームの面積を求めることができるからである。

【0056】

具体的には、図1のフレームページDP1は、行SQ1の記述から、（境界線L1によって）矢印D1方向に20%と80%（2対8）に分割されていることが分かり、領域Aを除いた領域は、行SQ2の記述から、（境界線L2とL3によって）矢印D2方向に30%と50%と20%（3対5対2）に分割されていることが分かり、これら領域A～Cを除いた領域は、行SQ3の記述から、（境界線L4によって）矢印D3方向に50%と50%（5対5）に分割されていることが分かるからである。

【0057】

フレーム構造の指定は、この%の替わりに、ピクセルなどで表現することも可能であるが、その場合でもフレーム規定ファイル中の前記行SQ1, SQ2, SQ3に相当する記述から、図1の各領域A～Eのフレームの面積を求めることができる点は同じである。

【0058】

なお、表示部24などの画面表示の絶対的な面積の値（面積値）は前記液晶ディスプレイの解像度などに大きく依存する（表示部24などの画面上では、解像度が高いほど小さく表示される）ため、例えば、絶対的な面積値（例えば、 $\text{cm}^2$ や $\text{mm}^2$ などを単位とする面積値）を求める場合には、解像度などの情報を予め当該領域抽出部101に与えておく必要がある。

【0059】

ただし本実施形態で面積値を求めるのは複数の領域のなかから最も大きな領域を選ぶためであるから、算出する面積値は必ずしも絶対的なものである必要はなく、1つのフレームページに関してのみ有効な相対的なものであれば十分である。相対的な面積値（各領域の面積の比率）を求める場合には、解像度などの情報は不要であり、領域抽出部101等の処理は簡単になる。

【0060】

当該領域抽出部101は、各領域（例えば、領域A～E）の面積の比率または



絶対的な面積値を算出すると、前記判定基礎情報として、所定の順番で最大領域判定部102へ供給する。

【0061】

この順番は、領域抽出部101と最大領域判定部102のあいだで整合が取れていればどのようなものであってもかまわないが、ここでは、一例として、画面上で上の領域ほど早く、同じ高さでは、左の領域ほど早く供給するものとする。したがって、図1に示したように画面表示されるフレームページDP1の場合なら、領域A、B、C、D、Eの順番で、各判定基礎情報が、最大領域判定部102に供給されることになる。

【0062】

最大領域判定部102は、受け取った判定基礎情報をもとに、1つのフレームページ中の主要領域を判定する部分である。

【0063】

各領域A～Eの内容を示すHTMLファイルDP12～DP16はその判定基礎情報とともに、当該最大領域判定部102へ供給するようにしてもよいが、HTMLファイルDP12～DP16の替わりに、HTMLファイルの識別情報を供給するようにしてもよい。当該識別情報としては、例えば、前記URL12～URL16（ファイル名）を利用することができる。

【0064】

識別情報を供給する場合、HTMLファイルDP12～DP16は、前記記憶部23内の例えば前記キャッシュ領域に格納しておき、主要領域を判定した後、最大領域判定部102が当該主要領域に対応するHTMLファイル（例えば、DP14）だけをキャッシュ領域から取り出すようにしてもよい。また、必要に応じて、最大領域判定部102は直接、HTMLファイルは取り扱わず、主要領域を判定した後には、その主要領域のHTMLファイルの識別情報だけを出力するようにしてもよい。

【0065】

その場合、その主要領域の画面表示を担当する画面表示モジュール（例えば、WWWブラウザB1）が、当該識別情報をもとに例えば前記キャッシュ領域から

、当該主要領域に対応するHTMLファイルを取り出してその画面表示を行うことになる。

【0066】

端子103は出力端子で、前記主要領域に対応するHTMLファイルまたはその識別情報を出力するために使用する。出力先は、当該画面表示モジュールである。

【0067】

一方、前記HTTPリクエストにこたえてHTTPレスポンスを返すWWWサーバ13の内部構成例は、図12に示す。

【0068】

(A-1-3) WWWサーバの内部構成例

図12において、当該WWWサーバ13は、通信部30と、制御部31と、記憶部32とを備えている。

【0069】

このうち通信部30は前記通信部20に対応し、制御部31は前記制御部21に対応し、記憶部32は前記記憶部23に対応するので、その詳しい説明は省略する。

【0070】

ただし制御部31は、WWWブラウザ(B1など)を搭載することはなくWWWサーバソフトを搭載している。

【0071】

また、WWWサーバ13が、予め生成したWWWページ等を蓄積しておくためのデータベースを伴う場合には、必要に応じて、DBMSも、当該制御部31に搭載され得る。

【0072】

さらに、前記フレームページDP1を提供するため、当該記憶部32には、少なくとも、HTMLファイルDP11~DP16が格納されている。

【0073】

以下、上記のような構成を有する本実施形態の動作について説明する。

【0074】

(A-2) 第1の実施形態の動作

ユーザU1が操作部22を操作することによって、または、前記オートパイロットツールなどの機能によって、URL11がWWWブラウザB1に入力されると、WWWブラウザB1が当該URL11に対応したHTTPリクエスト（GETリクエスト）を送信する。

【0075】

ただし前記領域処理部25に前記WWWブラウザB1と独立したHTTPクライアントとしての機能を持たせる場合には、この入力WWWブラウザB1ではなく、当該領域処理部25に対して行うことになる。

【0076】

通信端末12から送信されたこのHTTPリクエストをインターネット11経由で受信すると、WWWサーバ13（およびサーバOS）は、前記記憶部32から当該URL11によって指定されるフレーム規定ファイルDP11を取り出し、当該フレーム規定ファイルDP11をエンティティボディとして含むHTTPレスポンスを返送する。

【0077】

フレームページDP1を構成するその他のHTMLファイルDP12～DP14は、上述したように、当該フレーム規定ファイルDP11が通信端末12に受信されたあと、当該フレーム規定ファイルDP11中の記述（URL12～URL16）に基づいて、WWWブラウザB1が順次、自動的に送信する各HTTPリクエストに応じた各HTTPレスポンスのエンティティボディとして、通信端末12に受信されるから、これら（または、図3に示すように加工したデータ）を、前記領域処理部25が入力端子100を介して受け取ることになる。

【0078】

上述したように、判定基礎情報としての面積の比率（面積比率）や絶対的な面積値は、フレーム規定ファイル（ここでは、DP11）の記述だけから算出可能であるため、このとき領域処理部25内の領域抽出部101が受け取るのは、フレーム規定ファイルDP11だけでよく、HTMLファイルDP12～DP16

は受け取る必要はない。

【0079】

例えば、図3に対応する上述したフレーム規定ファイルDP11を受け取った場合、領域処理部25が当該ファイルDP11を処理した結果として出力するのは、図4に示す出力結果となる。

【0080】

図4中の各レコードRD1～RD5内の項目は、次のレコード構成(R1)の通りである。

【0081】

(分割領域の面積、分割領域のURL) … (R1)

このうち分割領域の面積は、上述した絶対的な面積値ではなく、フレームページDP1全画面の面積を100と置いた場合の、各領域の面積の比率を示している。

【0082】

具体的には、図3の行SQ1で全画面を20%と80%(2対8)に分割し、そのうちの20%が領域Aに対応するから、領域Aの面積比率は20(%)となる。同様に図3の行SQ2で、この領域Aの20%を除いた残りの80%を30%と50%と20%(3対5対2)に分割し、そのうちの30%が領域Bに対応し、50%が領域Cに対応するから、全画面における領域Bの面積比率は24( $=0.8 \times 0.3 \times 100$ )であり、全画面における領域Cの面積比率は40( $=0.8 \times 0.5 \times 100$ )である。

【0083】

領域D、Eに関しても同様の計算で、面積比率8(%)を求めることができる。

【0084】

また、当該レコード構成(R1)中で、分割領域のURLは、前記識別情報として、上述したURL12～URL16を含む記述をそのまま利用している。

【0085】

これらのレコードRD1～RD5を受け取った最大領域判定部102は、面積

比率を比較することにより、簡単に、面積比率40(%)の領域Cが最大であることを認識できるから、処理結果として、図5に示すように、当該領域Cに対応するHTMLファイルDP14のURLであるURL14を含む識別情報を出力することができる。

【0086】

この識別情報をもとに、WWWブラウザB1などの画面表示モジュールは、当該URL14で指定されるHTMLファイルDP14を例えば前記キャッシュ領域などから取得し、表示部24などに画面表示することが可能である。

【0087】

(A-3) 第1の実施形態の効果

以上のように、本実施形態によれば、領域が分割されて表示されるフレームページにおいて、ユーザ(例えば、U1)にとって最も重要な領域は最も面積の大きい領域であることが多いという性質を利用することで、自動的に主要領域を抽出することができる。

【0088】

これにより、ユーザ(例えば、U1)は、従来のように、開始箇所と終了箇所をあらかじめ手入力により指定する必要がなくなるから、ユーザの操作負担は、著しく軽減でき、ほとんど皆無である。

【0089】

したがって、例えば、多数のフレームページから主要領域だけを選択して出力させることも容易である。

【0090】

なお、本実施形態で、領域処理部(25)が実行する処理は、単なる面積値や面積比率の値の比較程度であるため、演算量は少なく、高速に結果を得ることができる。

【0091】

また、本実施形態は、基本的に自然言語処理を用いることなく実行可能であるため、記述言語(HTMLファイルの本文の内容など)に依存せずに主要領域を判定することができる。

【 0 0 9 2 】

さらに、本実施形態では、フレーム規定ファイルの解析を実施するが、予め指定した記述（例えば、前記行 S Q 1 ～ S Q 3， T G 1 2 ～ T G 1 6）だけ进行处理すればよいので、全ての記述を解析する場合などに比べ、処理量が少なく効率的である。

【 0 0 9 3 】

（ B ） 第 2 の 実 施 形 態

以下では、本実施形態が第 1 の実施形態と相違する点についてのみ説明する。

【 0 0 9 4 】

第 1 の実施形態では、面積（面積比率または絶対的な面積値）が大きな領域を主要領域であると判定したが、本実施形態では、フレームページ（例えば、 D P 1）の画面表示上の中心に近い位置に配置される領域を主要領域であると判定する。

【 0 0 9 5 】

（ B - 1 ） 第 2 の 実 施 形 態 の 構 成 お よ び 動 作

本実施形態と第 1 の実施形態は、前記領域処理部 2 5 の内部構成が相違するだけであるから、図 1 0，図 1 1，図 1 2 に示した構成は本実施形態でもそのまま用いることができる。本実施形態の領域処理部には符号 3 5 を付与して第 1 の実施形態の領域処理部 2 5 と区別する。

【 0 0 9 6 】

領域処理部 3 5 の内部構成例は図 6 に示す通りである。

【 0 0 9 7 】

図 6 において、当該領域処理部 3 5 は、入力端子 1 0 0 と、出力端子 1 0 3 と、領域抽出部 2 0 1 と、中心領域判定部 2 0 2 とを備えている。

【 0 0 9 8 】

このうち図 2 と同じ符号を付与した構成要素 1 0 0， 1 0 3 の機能は第 1 の実施形態と同じであるからその詳しい説明は省略する。

【 0 0 9 9 】

領域抽出部 2 0 1 は、第 1 の実施形態における領域抽出部 1 0 1 とは少し異な

り、入力されたフレーム規定ファイルDP11内に記述された領域分割に関する制御文字を解析することにより、分割された領域を抽出し、各領域の位置を示す位置情報を算出する部分である。

## 【0100】

各領域の位置の表現方法には様々なものがあり得るが、ここでは、一例として、図1に示す画面の左上端の点P0を原点とした場合の、矢印D1方向と矢印D2方向の座標で、各領域の2次元的な位置を表現する。また、矢印D1方向と矢印D2方向の座標の最小値は原点P0の0とし、最大値は100とする。

## 【0101】

各領域はすべて長方形であるから、例えば、その領域の左上端の点と、右下端の点の2次元的な位置（座標）を指定すれば、その領域の位置と大きさを特定することができ、間接的にはあるが、その領域の中心の位置も表現できる。このとき、2次元的な位置を表現する書式は、（D2方向の座標，D1方向の座標）であるものとする、各領域A～Eは、図7に示すレコードRD11～RD15によって表現できる。

## 【0102】

図7中の各レコードRD11～RD15内の項目は、次のレコード構成（R2）の通りである。

## 【0103】

（分割領域の左上端点の座標、分割領域の右下端点の座標、分割領域のURL） …（R2）

例えば領域Aの場合には、その左上端の点は、前記原点P0に一致し座標は（0，0）であり、右下端点P1の座標は（100，20）であるから、レコードRD11には、座標（0，0）、（100，20）が記述される。なお、このD1成分である20などは、図3に示した前記行SQ1などの記述をもとに算出することができる。

## 【0104】

その他の領域B～Eについてもこれと同様である。

## 【0105】

すなわち、領域Bの左上端点P2の座標は(0, 20)で右下端点P3の座標は(30, 100)であり、領域Cの左上端点P4の座標は(30, 20)で右下端点P5の座標は(80, 100)であり、領域Dの左上端点P6の座標は(80, 20)で右下端点P7の座標は(100, 60)であり、領域Eの左上端点P8の座標は(80, 60)で右下端点P9の座標は(100, 100)である。

## 【0106】

したがって、レコードRD12～RD15内には、それぞれ該当する左上端点および右下端点の座標が記述される。

## 【0107】

領域抽出部35は、図3に示したデータ(フレーム規定ファイルDP11の主要部)をもとに、これらレコードRD11～RD15を生成し、中心領域判定部202へ供給する。各レコードRD11～RD15を供給するときの順番は、第1の実施形態と同様、領域A、B、C、D、Eの順番、すなわち、レコードRD11、RD12、RD13、RD14、RD15の順番であってよい。

## 【0108】

これらの供給を受ける中心領域判定部202は、レコードRD11～RD15を処理し、領域の中心が、表示部24の画面の中心(フレームページの中心)CPに最も近い領域を、主要領域と判定する部分である。

## 【0109】

なお、WWWブラウザB1のウィンドウ(ブラウザウィンドウ)を小さく表示させた場合などには、表示部24(液晶ディスプレイなど)の画面の中心とブラウザウィンドウの中心(またはフレームページの中心)とは一致しないから、表示部24の画面の中心よりも、フレームページ(ここでは、DP1)の中心またはブラウザウィンドウの中心というほうが、より本質的である。

## 【0110】

上述したように原点P1の座標を(0, 0)とし、点P9の座標を(100, 100)とした場合には、フレームページDP1の中心CPの座標は(50, 50)となるから、中心領域判定部202は、各領域の中心と当該フレームページ



の中心CPとの距離を求めることになる。

#### 【0111】

なお、距離の単位は、cm、mm、ピクセル数など、どのようなものを用いてもかまわないが、本実施形態では、このような単位を用いて距離を絶対的に表現せず、相対的な距離の比率を利用している。上述した面積比率の場合と同様、相対的な距離の比率だけを用いることで、中心領域判定部202等の処理を簡単にすることができる。

#### 【0112】

各レコードRD11～RD15には、各領域の左上端点と右下端点の座標が記述されているから、各領域の中心の座標は、左上端点と右下端点のあいだでD2成分とD1成分の平均値を取ることによって容易に算出可能である。

#### 【0113】

例えば、前記領域Aの中心の位置は、左上端点P0(0, 0)と右下端点P1(100, 20)のD2成分とD1成分の平均値を取って、(50, 10)となる。

#### 【0114】

同様に、例えば、領域Cの中心の位置は、その左上端点P4(30, 20)と右下端点P5(80, 100)のD2成分とD1成分の平均値を取って、(55, 60)となる。

#### 【0115】

他の領域B、D、Eについても同様な処理で、その中心の座標を求めることができる。

#### 【0116】

図1の例の場合、前記座標(55, 60)がフレームページDP1の中心CP(50, 50)に最も近いから、中心領域判定部202は、当該座標(55, 60)に対応する領域Cを主要領域と判定することになる。したがって、中心領域判定部202は、領域Cに対応するHTMLファイルDP14のURLであるURL14を、出力端子103から出力することになる。

#### 【0117】

(B-3) 第2の実施形態の効果

本実施形態では第1の実施形態の効果とほぼ同等な効果を得ることができる。

【0118】

ただし第1の実施形態では、面積（面積比率または絶対的な面積値）が大きな領域を主要領域であると判定したのに対し、本実施形態では、フレームページ（例えば、DP1）の画面表示上の中心に近い位置に配置される領域を主要領域であると判定するため、フレームページのフレーム構造によっては、第1の実施形態とは異なる領域を主要領域と判定する可能性がある。

【0119】

(C) 第3の実施形態

以下では、本実施形態が第1、第2の実施形態と相違する点についてのみ説明する。

【0120】

本実施形態は第1の実施形態と第2の実施形態を複合したものである。

【0121】

(C-1) 第3の実施形態の構成および動作

本実施形態と第1の実施形態は、前記領域処理部25の内部構成が相違するだけであるから、図10、図11、図12に示した構成は本実施形態でもそのまま用いることができる。本実施形態の領域処理部には符号45を付与して第1の実施形態の領域処理部25や第2の実施形態の領域処理部35と区別する。

【0122】

領域処理部45の内部構成例は図8に示す通りである。

【0123】

図8において、当該領域処理部45は、入力端子100と、出力端子103と、領域抽出部301と、最大領域判定部102Aと、中心領域判定部202Aと、選択部303とを備えている。

【0124】

このうち図2と同じ符号を付与した構成要素100、103の機能は第1の実施形態と同じであるからその詳しい説明は省略する。

## 【0125】

領域抽出部301は、第1の実施形態の領域抽出部101と第2の実施形態の領域抽出部201の機能を兼ね備えた部分で、例えば、図3のデータを受け取ったときには、図4および図7のレコードRD1～RD5、RD11～RD15を出力する。レコードRD1～RD5、RD11～RD15の替わりに、これらをまとめた、図9に示すレコードRD21～RD25を出力することも望ましい。

## 【0126】

図9中の各レコードRD21～RD25内の項目は、次のレコード構成(3)の通りである。

## 【0127】

(分割領域の面積、分割領域の左上端点の座標、分割領域の右下端点の座標、分割領域のURL) … (3)

最大領域判定部102AにはこれらレコードRD1～RD5とRD11～RD15が供給され、中心領域判定部202Aにも、レコードRD1～RD5とRD11～RD15が供給される。あるいは、これらをまとめた内容を有するレコードRD21～RD25が両判定部102A、202Aに供給される。

## 【0128】

最大領域判定部102Aは基本的に第1の実施形態の最大領域判定部102と同じ機能を持つ部分であるが、領域の面積(面積比率でも、絶対的な面積値でもかまわない。この点は、以下の説明でも同じ。)に関する処理と無関係なレコードRD11～RD15を選択部303へ渡す機能を有する点が相違する。レコードRD11～RD15を選択部303へ渡すまえに、面積が最大の領域(例えば、領域C)に対応するレコード(例えばRD13)だけを選択し、選択したレコードだけを、面積が最大の領域に対応するHTMLファイルのURLとともに選択部303へ供給することは、選択部303の処理量を軽減する上で望ましい。

## 【0129】

具体的な実装にも依存するが、供給を受けたレコードがRD21～RD25である場合などには、当該最大領域判定部102A内で、実質的に、この選択を行うほうが効率的であると考えられる。

## 【 0 1 3 0 】

同様に、中心領域判定部 2 0 2 A は基本的に第 2 の実施形態の中心領域判定部 2 0 2 と同じ機能を持つ部分であるが、中心間の距離に関する処理と無関係なレコード R D 1 ~ R D 5 を選択部 3 0 3 へ渡す機能を有する点が相違する。レコード R D 1 ~ R D 5 を選択部 3 0 3 へ渡すまえに、フレームページの中心に最も中心の位置に近い領域（例えば、領域 C）に対応するレコード（例えば R D 3）だけを選択し、選択したレコードだけを、フレームページの中心に中心が最も近い領域に対応する HTML ファイルの URL とともに選択部 3 0 3 へ供給することは、選択部 3 0 3 の処理量を軽減する上で望ましい。

## 【 0 1 3 1 】

具体的な実装にも依存するが、供給を受けたレコードが R D 2 1 ~ R D 2 5 である場合などには、当該中心領域判定部 2 0 2 A 内で、実質的に、この選択を行うほうが効率的であると考えられる。

## 【 0 1 3 2 】

最大領域判定部 1 0 2 A と中心領域判定部 2 0 2 A から供給されるデータをもとに、選択部 3 0 3 は、次の式（F 1）に応じた選択処理を実行する。

## 【 0 1 3 3 】

$$V = X \alpha + Y \beta \quad \cdots (F 1)$$

ここで、X は面積、Y はフレームページの中心と領域の中心との距離の逆数を表す。また、 $\alpha$ 、 $\beta$  は、重み係数である。当該  $\alpha$ 、 $\beta$  は、希望通りの結果が得られやすいように、ユーザ U 1 が適宜、設定できるようにすることも望ましい。

## 【 0 1 3 4 】

この選択処理では、最大領域判定部 1 0 2 A から入力されるデータから式（F 1）で計算される値（V 1 とする）と、中心領域判定部 2 0 2 A から入力されるデータから式（F 1）で計算される値（V 2 とする）を比較し、V 1 と V 2 のうちいずれか大きいほうのデータを選択し、選択されたデータに対応する領域の HTML ファイルの識別情報（URL）を、出力端子 1 0 3 から出力する。

## 【 0 1 3 5 】

式（F 1）に基づく選択処理が必要とされるのは、判定部 1 0 2 A による判定

結果と判定部202Aによる判定結果が異なるケースであるから、両判定結果が一致する場合には、式(F1)による選択処理は省略するようにしてもよい。

【0136】

(C-2) 第3の実施形態の効果

本実施形態によれば、第1または第2の実施形態と同等な効果を得ることができる。

【0137】

加えて、ユーザ(U1)が式(F1)の重み係数( $\alpha$ 、 $\beta$ )の設定を行うことができるように構成した場合には、第1の実施形態と第2の実施形態で判定結果が相違し得るケースでも、柔軟に、ユーザの希望に適合した主要領域を出力することが可能である。

【0138】

(D) 他の実施形態

上記第3の実施形態では、前記式(F1)を用いて選択処理を行ったが、本発明では、式(F1)以外の式をもとに選択処理を行うようにしてもかまわない。

【0139】

一例として、次の式(F2)をもとに選択処理を行うことも可能である。

【0140】

$$V = XY \quad \dots (F2)$$

このX、Yの意味は前述した通りである。

【0141】

また、上記第3の実施形態では、選択部303の選択処理として、面積と位置情報から計算された値を用いているが、別の情報を用いて選択処理を行うことも可能である。例えば、領域に表示される同じ文字種で同サイズの文字の数が多いほうを選択するようにしてもよい。

【0142】

また、上記第1～第3の各実施形態において、領域抽出部が各領域に関する判定結果を出力する順番は、上述したものに限らないことは当然である。例えば、順次、右の分割領域から左へ、上の分割領域から下へと出力するラスタースキャ

ンの順序としてもよく、その他の順序（ラスタースキャンと逆の順序などとしてもよい。

【0143】

なお、上記第1、第3の実施形態では、スクロール分を無視してフレームの面積や、フレームの面積の比率をもとに主要領域を判定したが、必要に応じて、スクロール分も加味し画面表示されるコンテンツ（HTMLファイル）の面積比率や絶対的な面積値を算出するようにしてもよい。

【0144】

あるHTMLファイルの画面表示されたときの面積がそのHTMLファイルが配置されるフレームよりも大きいと自動的にスクロールバーが表示され、縦方向（例えば、前記矢印D1方向）または横方向（例えば、前記矢印D2方向）にスクロールすることができるようになる。縦方向に大きい場合には、縦方向にスクロールするための縦スクロールバーが表示され、横方向に大きい場合には、横方向にスクロールするための横スクロールバーが表示される。

【0145】

スクロール分を加味した面積比率や絶対的な面積値を求める場合、概略的な指標として、当該HTMLファイルのファイルサイズを利用することができる。直接的には画面表示の対象とならないタグなどの制御文字の記述が多ければ、画面表示される面積値は変わらなくてもHTMLファイルのファイルサイズは増大し、またファイルサイズが同じでもフォントの指定などによってスクロール分を含む画面表示面積が増大すること等もあるため、ファイルサイズは厳密には表示の面積値や面積の比率に対応しているものではないが、画面に表示される文字などの記述と表示されない制御文字などの記述の割合がHTMLファイル間でそれほど大きく相違しない場合や、フォントの指定の頻度や程度がHTMLファイル間でそれほど大きく相違しない場合などには、前記ファイルサイズは、面積比率や絶対的な面積値を示す概略的な指標として良好なものとなる。

【0146】

厳密には、HTMLファイル中のすべてのタグを解釈して適切な演算を行えば、スクロール分も含めて、面積比率や絶対的な面積値を高精度に算出することも

可能であるが、それでは演算量が大きくなってしまうため、ファイルサイズを用いるのが簡便である。

【0147】

ファイルサイズは、前記HTTPレスポンスに含まれるHTTPヘッダのうち、エンティティヘッダに記述される情報のなかの1つであるため、容易に取得可能である。

【0148】

なお、当該エンティティボディには、ファイル形式に関する情報も含まれているため、フレームページを構成するいずれかのHTMLファイル（例えば、DP12～DP16）に画像ファイルなどがその領域の一部としてリンクされていても、当該画像ファイルなどは除外してHTMLファイルだけ選別し、HTMLファイルのファイルサイズだけを比較すること等が可能である。

【0149】

また、上記第1～第3の実施形態において、各領域の絶対的な面積値を求める場合には、表示部24などに関係なく、予め標準的な環境に対応した値（解像度など）を与えておけば、表示部24に対応するディスプレイ装置などを他の仕様のディスプレイ装置に置き換えた場合にも、その解像度などを前記領域処理部25に供給する必要がなく、簡便である。

【0150】

さらに、前記領域処理部25、35、45は、自動的にユーザが必要とする領域を特定できるため、（イ）指定ウェブページのうち必要箇所が更新されたときにユーザに通知する通知システムや、（ロ）必要箇所（主要領域）以外は検索対象としない等の機能を持つ効率的な検索システム、（ハ）必要箇所のみを要約対象とする効率的な要約システムなどの一部として利用しても有用である。

【0151】

また、フレームページの場合、図1に示した領域Bのような狭いフレームにメニューを置き、領域Cのような広いフレームにそのメニューの選択に応じて変わる内容を置く構成が典型例であるが、メニューのないフレームページにも、本発明が適用できることは当然である。

【 0 1 5 2 】

さらに、フレームページ以外のWWWページに対しても本発明は適用可能であり、HTML以外の言語（XMLやSGMLなど）による記述にも適用可能である。何らかの意味で、論理的に識別可能な複数の領域が含まれていればよいからである。

【 0 1 5 3 】

また、ファイル以外の単位をもとに領域を識別してもよいことは当然である。

【 0 1 5 4 】

さらに、使用する通信プロトコルは必ずしもHTTPでなくともかまわない。

【 0 1 5 5 】

なお、上記第1、第2の実施形態では通信端末（クライアント）12側に領域処理部25、35、45を配置したが、当該領域処理部の機能は、WWWサーバ13側に配置したり、WWWサーバ13と通信端末12の中間に介在し得る例えばプロキシサーバなどに配置することも可能である。

【 0 1 5 6 】

WWWサーバ13側に配置した場合には、必ずしもHTTPによる通信を行わなくてもよい。

【 0 1 5 7 】

また、上記第1～第3の実施形態では、フレームページがWWWサーバ13上に公開されたものであることを前提としているが、CD-ROMなどの記録媒体から得たフレームページ等にも本発明は適用できるので、対象とするフレームページ等は、必ずしもネットワーク経由で入手されるものでなくともかまわない。

【 0 1 5 8 】

以上の説明では主としてハードウェア的に本発明を実現したが、本発明はソフトウェア的に実現することも可能である。

【 0 1 5 9 】

【発明の効果】

本発明によれば、構造化文書の中から主要な領域を選出するために必要なユーザの操作負担を軽減することができる。



【図面の簡単な説明】

【図 1】

WWW文書の構成例の説明図である。

【図 2】

第 1 の実施形態の領域処理部の機能的構成を示すブロック図である。

【図 3】

第 1 の実施形態の領域処理部への入力データ例を示す説明図である。

【図 4】

第 1 の実施形態の領域抽出部による抽出結果を示す説明図である。

【図 5】

第 1 の実施形態の領域処理部の出力結果例を示す説明図である。

【図 6】

第 2 の実施形態の領域処理部の機能的構成を示すブロック図である。

【図 7】

第 2 の実施形態の領域抽出部による抽出結果を示す説明図である。

【図 8】

第 3 の実施形態の領域処理部の機能的構成を示すブロック図である。

【図 9】

第 3 の実施形態の領域抽出部による抽出結果を示す説明図である。

【図 10】

第 1 および第 2 の実施形態の通信システムの全体構成例を示す概略図である。

【図 11】

第 1 および第 2 の実施形態で使用する通信端末の構成例を示す概略図である。

【図 12】

第 1 および第 2 の実施形態で使用する WWW サーバの構成例を示す概略図である。

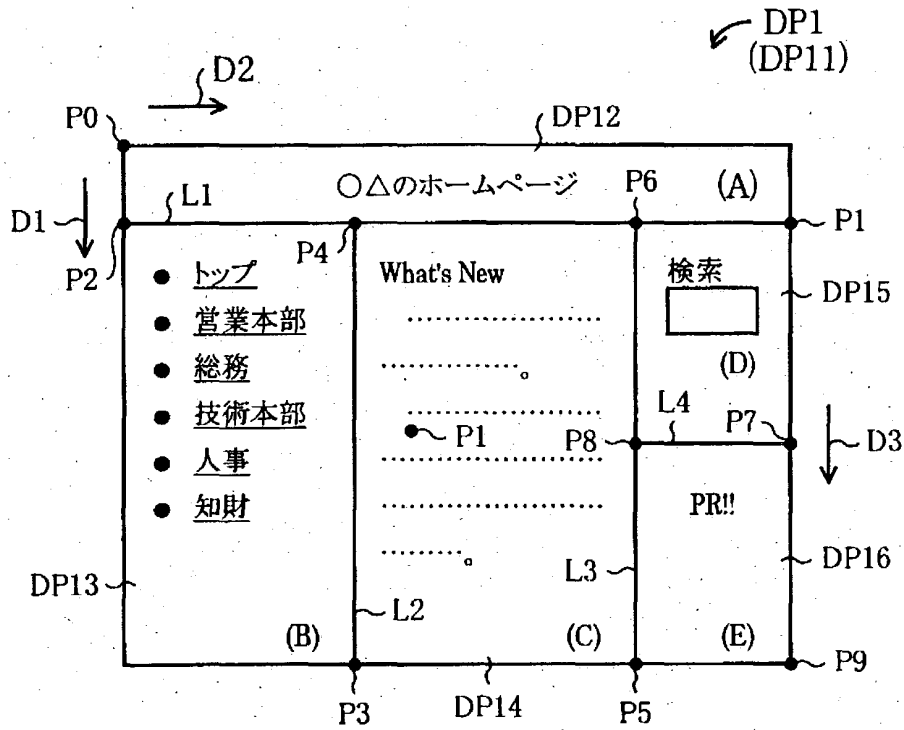
【符号の説明】

10…通信システム、11…ネットワーク、12…通信端末、13…WWWサーバ、20、30…通信部、21、31…制御部、22…操作部、23、32…

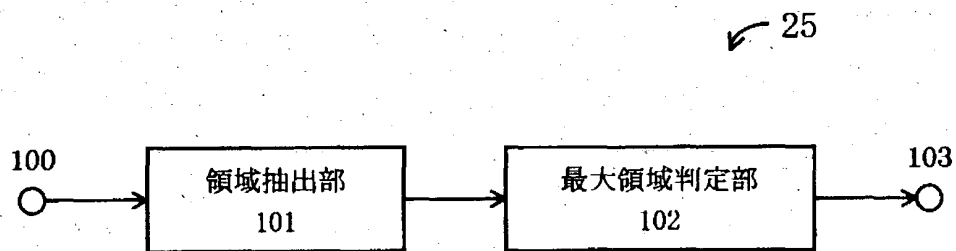
記憶部、2 4 …表示部、2 5, 3 5 …領域処理部、1 0 1、2 0 1、3 0 1 …領域抽出部、1 0 2、1 0 2 A …最大領域判定部、2 0 2、2 0 2 A …中心領域判定部、3 0 3 …選択部。

【書類名】 図面

【図 1】



【図 2】



【図3】

DP11

```

<HTML>
SQ1  <FRAMESET ROWS="20%,80%">
TG12 <FRAME SRC="title.html">    ←(A)の部分のソース
SQ2  <FRAMESET COLS="30%,50%,20%">
TG13 <FRAME SRC="link.html">    ←(B)の部分のソース
TG14 <FRAME SRC="honbun.html">  ←(C)の部分のソース
SQ3  <FRAMESET ROWS="50%,50%">
TG15 <FRAME SRC="sonota1.html"> ←(D)の部分のソース
TG16 <FRAME SRC="sonota2.html"> ←(E)の部分のソース
    </FRAMESET>
    </FRAMESET>
    </FRAMESET>
</HTML>
    
```

PS1

【図4】

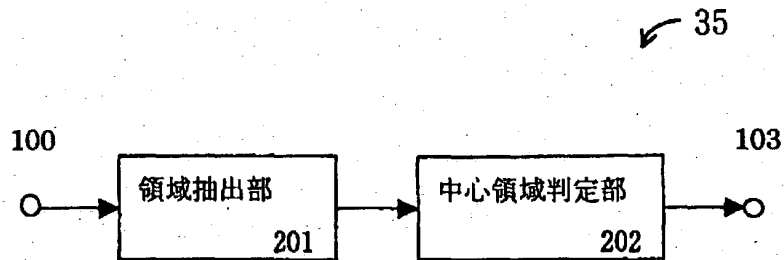
```

RD1  (20, <FRAME SRC="title.html">)    ←(A)の部分
RD2  (24, <FRAME SRC="link.html">)      ←(B)の部分
RD3  (40, <FRAME SRC="honbun.html">)    ←(C)の部分
RD4  (8, <FRAME SRC="sonota1.html">)    ←(D)の部分
RD5  (8, <FRAME SRC="sonota2.html">)    ←(E)の部分
    
```

【図 5】

```
<FRAME SRC="honbun.html">
```

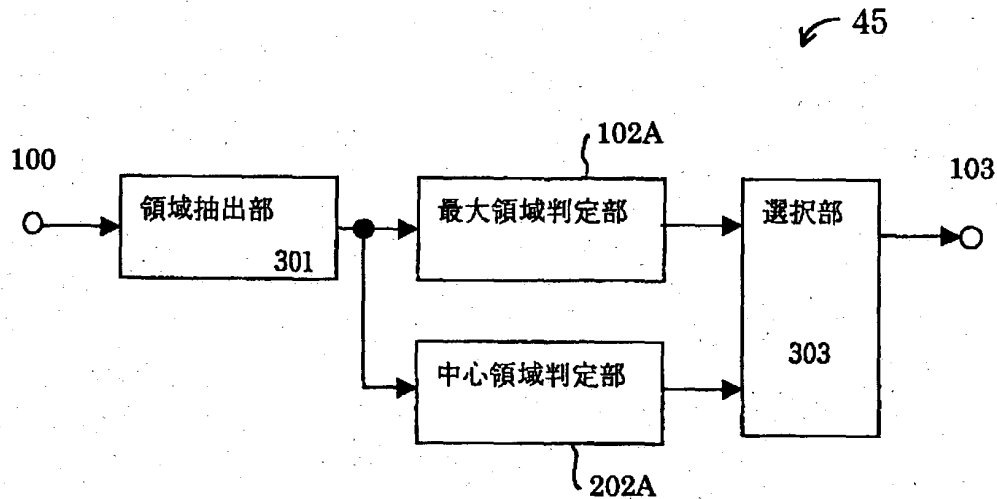
【図 6】



【図 7】

RD11	←	( (0,0), (100,20), <FRAME SRC="title.html">)	←(A)の部分
RD12	←	( (0,20), (30,100), <FRAME SRC="link.html">)	←(B)の部分
RD13	←	( (30,20), (80,100), <FRAME SRC="honbun.html">)	←(C)の部分
RD14	←	( (80,20), (100,60), <FRAME SRC="sonota1.html">)	←(D)の部分
RD15	←	( (80,60), (100,100), <FRAME SRC="sonota2.html">)	←(E)の部分

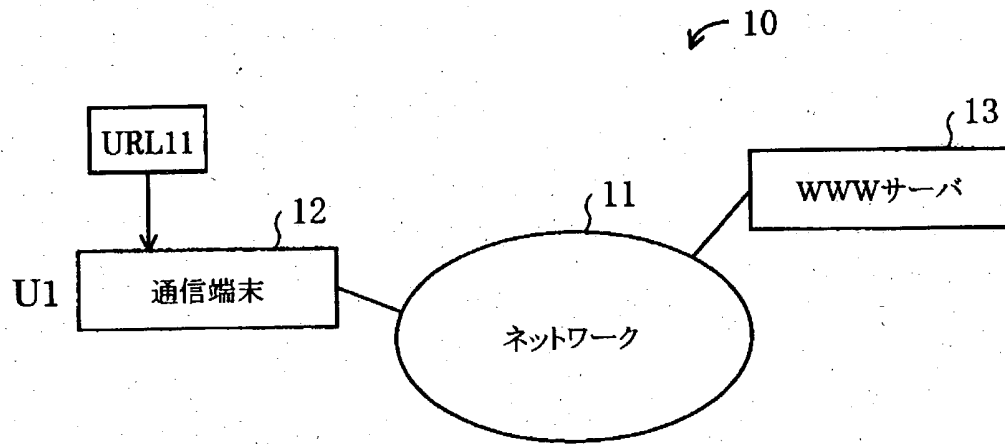
【図 8】



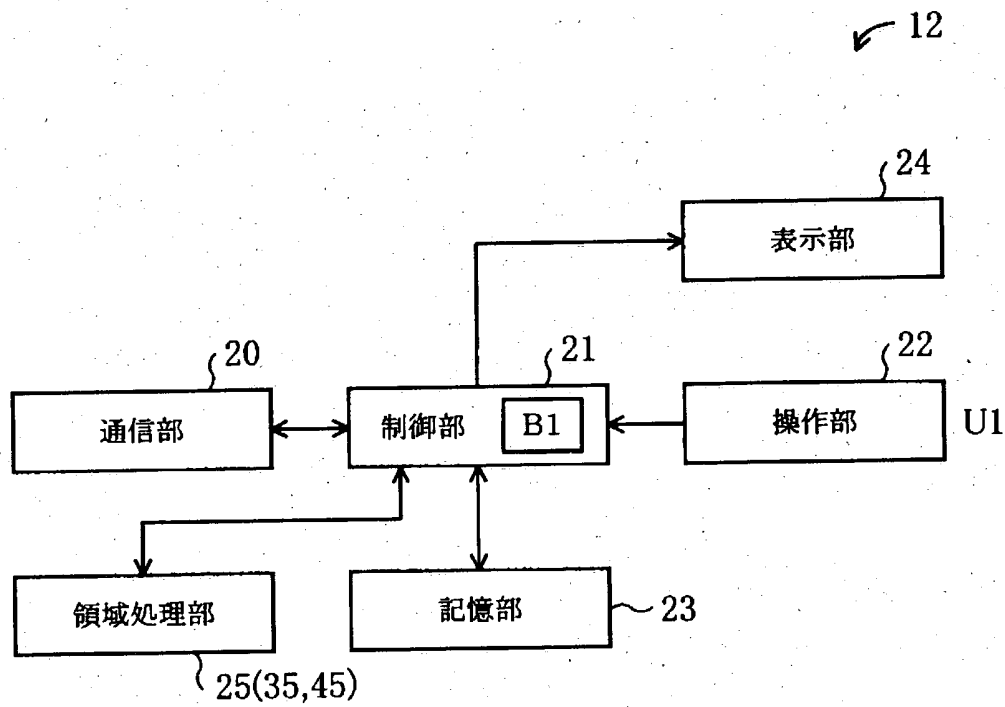
【図 9】

RD21	(20, (0,0), (100,20), <FRAME SRC="title.html">)	←(A)の部分
RD22	(24, (0,20), (30,100), <FRAME SRC="link.html">)	←(B)の部分
RD23	(40, (30,20), (80,100), <FRAME SRC="honbun.html">)	←(C)の部分
RD24	(8, (80,20), (100,60), <FRAME SRC="sonota1.html">)	←(D)の部分
RD25	(8, (80,60), (100,100), <FRAME SRC="sonota2.html">)	←(E)の部分

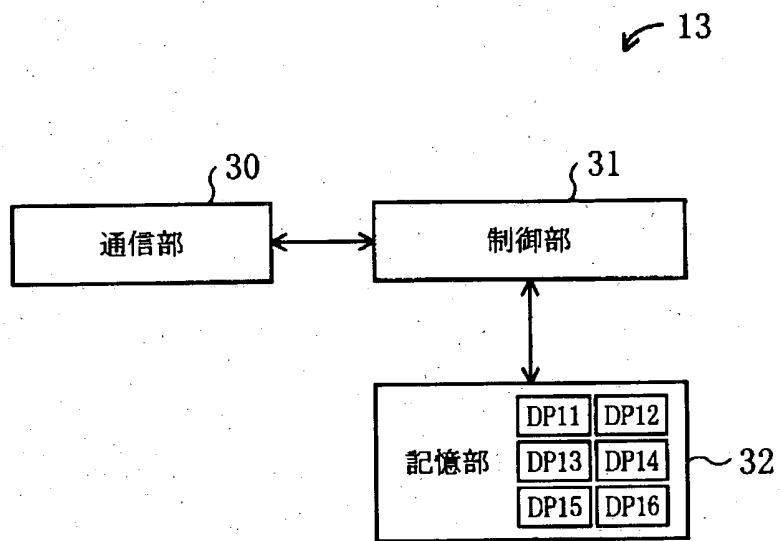
【図10】



【図11】



【図 12】





【書類名】 要約書

【要約】

【課題】 操作負担を軽減する。

【解決手段】 複数の領域を含む所定の構造化文書の中から、主要な領域を選出する情報処理装置において、前記構造化文書の表示構造を指定する制御文字を解析し、前記各領域に関する幾何学的な属性情報を生成する属性情報生成部と、前記属性情報を比較することによって複数の領域の中から主要領域を選出する主要領域選出部とを備える。このような構成において、前記属性情報生成部は、前記属性情報として各領域の表示面積または表示面積の比率を用い、前記主要領域選出部は、当該表示面積または表示面積の比率が大きい領域を主要領域として選出することも好ましい。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [000000295]

1. 変更年月日 1990年 8月22日

[変更理由] 新規登録

住 所 東京都港区虎ノ門1丁目7番12号

氏 名 沖電気工業株式会社